Combining soil spectral reflectance data and satellite imagery to assess impacts of land use on soil fertility in Tajikistan

Wolfgramm, Bettina ; Seiler, Bruno ; Guntli, David ; Liniger, Hanspeter ; Shepherd, Keith ; Kneubühler, Mathias ; Kellenberger, Tobias

Abstract: Recent studies showed that soil fertility properties can be predicted from soil spectral reflectance data and in a second step can be combined successfully with information from satellite imagery for rapid assessment of soil quality over large areas. This approach shall be adapted for a test area in the Loess zone of Tajikistan in order to assess the impact of land use on soil fertility. The groundtruth data collected confirms that widespread land use changes have taken place since 1992 (30 % of the area formerly used as grazing land has been cultivated since 1992). The newly cultivated areas are situated on steep slopes (the average slope is 20 %) and show visible signs of water erosion in 60 % of the cases observed. Also 48 % of the plots recorded from grazing land showed signs of water erosion. VIS-NIR measurements of soil samples collected from each sampling plot have been explored for relations between soil reflectance data and commonly used indicators of soil fertility in the study area. First results show that reflectance wavebands are strongly relating to CaCO3 and soil colour. Regression tree modelling has been carried out successfully to calibrate total nitrogen contents determined by chemical analysis against reflectance wavebands (validation r2 for regression was 0.71). A classification tree model predicting areas with water erosion shows the potential of decision tree modelling when combining different datasets. Hierarchical structures can be revealed and thresholds for mapping purposes using raster datasets available (DEM and Landsat 7 satellite imagery) can be determined. Prediction success determined by 10 fold cross-validation was 72 % and 61% for the classes erosion and no erosion respectively.

Originally published at:
Wolfgramm, Bettina; Seiler, Bruno; Guntli, David; Liniger, Hanspeter; Shepherd, Keith; Kneubühler, Mathias; Kellenberger, Tobias (2005). Combining soil spectral reflectance data and satellite imagery to assess impacts of land use on soil fertility in Tajikistan. In: 1st International Conference on Remote Sensing and Geoinformation Processing in the Assessment and Monitoring of Land Degradation and Desertification (RGLDD), Trier, Germany, 7 September 2005 - 9 September 2005. s.n., 611-617.

Combining soil spectral reflectance data and satellite imagery to assess impacts of land use on soil fertility in Tajikistan

Bettina Wolfgramm[a], Bruno Seiler[b], David Guntli[b], Hanspeter Liniger[a]

Keith Shepherd[c], Mathias Kneubühler[b], and Tobias Kellenberger[b]

[a] Centre for Development and Environment, Institute of Geography
University of Berne, Steigerhubelstrasse 3, 3008 Berne
Switzerland, email: Bettina.Wolfgramm@cde.unibe.ch

[b] Remote Sensing Laboratories (RSL), University of Zurich
Winterthurerstrasse 190, 8057 Zurich, Switzerland

[c] World Agroforestry Centre (ICRAF), P.O. Box 30677-00100, Nairobi, Kenya

ABSTRACT

Recent studies showed that soil fertility properties can be predicted from soil spectral reflectance data and in a second step can be combined successfully with information from satellite imagery for rapid assessment of soil quality over large areas. This approach shall be adapted for a test area in the Loess zone of Tajikistan in order to assess the impact of land use on soil fertility. The groundtruth data collected confirms that widespread land use changes have taken place since 1992 (30 % of the area formerly used as grazing land has been cultivated since 1992). The newly cultivated areas are situated on steep slopes (the average slope is 20 %) and show visible signs of water erosion in 60 % of the cases observed. Also 48 % of the plots recorded from grazing land showed signs of water erosion. VIS-NIR measurements of soil samples collected from each sampling plot have been explored for relations between soil reflectance data and commonly used indicators of soil fertility in the study area. First results show that reflectance wavebands are strongly relating to $CaCO_3$ and soil colour. Regression tree modelling has been carried out successfully to calibrate total nitrogen contents determined by chemical analysis against reflectance wavebands (validation r^2 for regression was 0.71). A classification tree model predicting areas with water erosion shows the potential of decision tree modelling when combining different datasets. Hierarchical structures can be revealed and thresholds for mapping purposes using raster datasets available (DEM and Landsat 7 satellite imagery) can be determined. Prediction success determined by 10 fold cross-validation was 72 % and 61 % for the classes erosion and no erosion respectively.

Keywords: diffuse reflectance spectroscopy, land cover / land use, land degradation / conservation assessment, soil fertility, Tajikistan.

1 INTRODUCTION

In the 1990's increasing poverty triggered by the civil war and the transformation of the economy lead to widespread cultivation of steep slopes formerly used as grazing land. The starting point for this development, was March 1992 when the Presidential Decree "On Renting Land" authorized collectives ore state farms to rent and lease land to households for independent cultivation [1]. The foothills of western Tajikistan consist mainly of easily erodable Loess deposits. Water erosion is considered to be the fastest and most widespread degradation process [2]. Examples of farmers searching for alternative land management types with soil and water conservation (SWC) measures such as area closure by fencing, where vegetation cover has been increased several times, give an idea of the potential of the area with regard to sustainable land management [3].

Today governmental offices in Tajikistan lack the resources for monitoring land degradation at the national level and up-to-date information on the state of natural resources in the rainfed areas is not available. Cost and time efficient as well as reliable methods and combinations of methods for generating up-to-date information for assessments over large areas are needed.

Presented at the RGLDD International Conference, Trier (Germany), September 7th-9th, 2005

The spectral library approach recently developed at the World Agroforestry Centre (ICRAF) in Kenya is designed for rapid determination of soil fertility properties for big numbers of samples, which is opening up opportunities for risk assessments. VIS-NIR measurements of air-dried soil samples carried out under standardized conditions in the laboratory have been calibrated to soil properties determined by chemical analysis for a wide range of African soils [4]. Further it has been shown that soil fertility indexes integrating commonly used agronomic indicators of soil fertility can be developed from reflectance spectral data and be calibrated to local conditions, allowing the spatial representation of soil fertility based on remote sensing satellite imagery [5].

This approach shall be adapted for a test area in the Loess zone of Tajikistan in order to assess the impact of land use on soil fertility. The objectives of this paper are: (i) to give an overview on crucial land cover/land use change and land degradation issues based on groundtruth data collected, (ii) to take first steps in adapting the spectral library approach to predict soil properties from spectral measurements to Tajik conditions and (iii) to test opportunities for mapping land cover/land use and land degradation/conservation classes by combining various datasets using classification tree modelling.

2 MATERIALS AND METHODS

2.1 The study area – general description

The three test areas of this study are all situated on the Loamy Loess and are defined as calcareous mountain cinnamonic soils by the local Tajik definition system. Typical values for soil particle size distribution are 5 % sand, 60 % silt and 25 % clay. $CaCO_3$ contents vary between 2 to 30 %, depending on the mother rock, but also on the state of erosion. In the topsoil average contents of TN are 0.15-0.25 % and soil organic carbon (SOC) 1-2 % [6]. Rainfall characteristics vary from the South with 400 mm per year to the Northeast with up to 900 mm. Rainfall distribution is alike in all the area and rains are concentrated during November through to April. The main crop is winter wheat. The fields are prepared (ploughed or harrowed) in November before the rains start. In rotation (every 2-4 years) flax, chickpeas and beans are planted. Since tractors cannot be driven along the contours of the steep slopes, fields are ploughed up and down, wherever tractors are available.

2.2 Overall study approach

Figure 1 gives an overview on groundtruth collected, materials used, methods applied and first results achieved that will be presented in this paper.

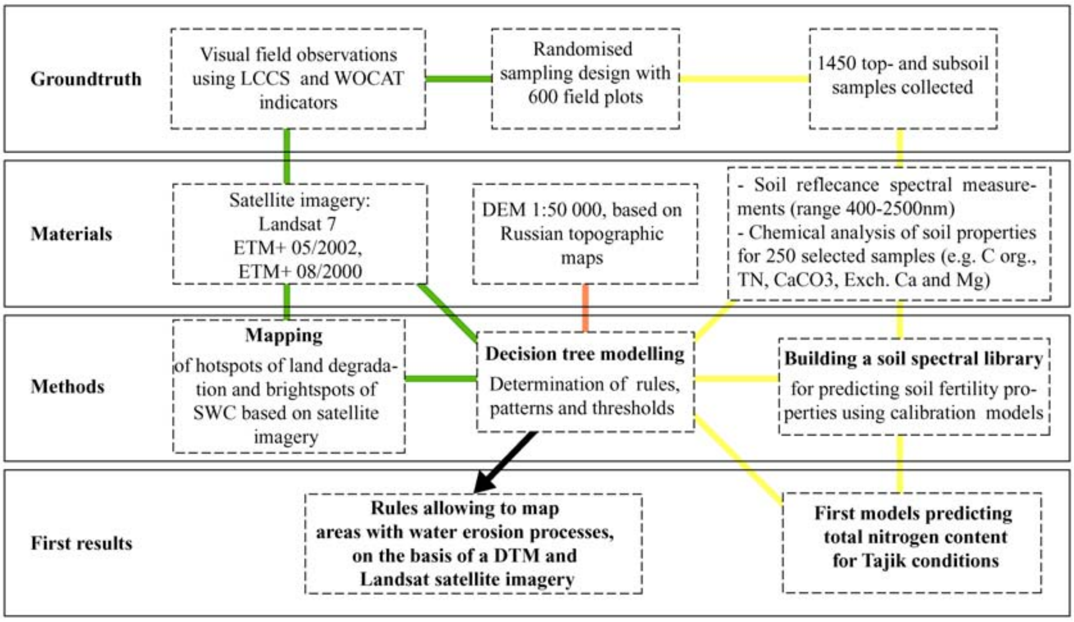

Figure 1. Overall approach

2.3 Raster data sets

The bases for mapping and analysis provide a digital elevation model (DEM) calculated from Russian topographic maps (1:50 000, 20 m equidistance) and Landsat 7 (ETM+) imagery from various years and seasons. Satellite imagery rectification was performed using GPS Ground Control Points measured in the field and additional control points extracted from Russian topographic maps. Differential GPS was not available. Residuals in x direction were for both Landsat 7 images on average 53 m, y-residuals amounted to 11 m for the image from 2002 and 20 m for the one from 2000. The atmospheric correction of the georeferenced images was conducted using ATCOR3 [7]. The main focus so far has been on the land cover/land use situation and the state of the soil in the years 2000. The state of fullest development of vegetation is represented by an image dating from 24.05.2002 and the situation after harvest, during the dry season and best visibility of bare ground is represented by an image dating from 22.08.2000. Given the importance of vegetation characteristics when assessing land degradation, the Optimised Soil Adjusted Vegetation Index (OSAVI) [8] was derived from the Landsat 7 satellite imagery. By adding the soil-adjustment parameter 0.16 it is attempted to minimize brightness-related soil effects. Surface reflectance for each sampling point was extracted from the Landsat 7 scene and from the DEM and served as input into land cover/land use and land degradation models.

2.4 Groundtruth

Test areas of 10x10km were selected based on a first assessment of seasonal changes in vegetation, approximated by information derived from the calculated OSAVI layers. This assessment based on the assumption that (i) if during the main vegetation period areas are showing no vegetation, it is a sign of severe degradation (hotspots) and (ii) on the other hand if after harvesting of rainfed crops, during the very dry months of July to October areas show dense vegetation this land use type is likely to conserve the land resources (brightspots). Test areas were placed in regions showing similar distributions of hot- and brightspots and representing different ecological zones.

A randomised sampling design as described by Walsh [9] was chosen for full characterization of the area, allowing geo-statistical and risk assessments. It included 600 plots, clustered in groups of 13 plots. These clusters were again distributed over the 3 test areas, each including 15 clusters. Plots size was approximately 30 by 30 m, corresponding with the pixel size of Landsat 7 scenes.

Groundtruth was collected for visible indicators of land degradation and conservation according to the World Overview on Conservation Approaches and Technologies (WOCAT) [10],[11] and land cover classes according to the FAO land cover classification system [12]. According to WOCAT land degradation characteristics the field protocol included water erosion, sediment deposition, crusting, cracks, compaction and mass movements. Conservation classes recorded were (typical examples from Tajikistan for each conservation class are given in brackets): Agronomic (e.g. contour ploughing), vegetative (e.g. grass strips, windbreaks consisting of poplar trees), structural (e.g. terraces introduced in Soviet times) and management measures (e.g. private area closure by fencing). Land cultivation history types have been determined in the field based on visible observations such as indicator plants, signs of ploughing and former field boundaries. The information noted down in the field was then compared or crosschecked against land use maps at a scale 1:10 000 dating from 1985. On each sampling plot topsoil (0-20 cm) and subsoil (20-50 cm) samples were collected as composite samples from two sampling pits. Soil chemical and physical analysis are being conducted at the laboratories of the World Agroforestry Centre (ICRAF) in Kenya using standard laboratory procedures. The following soil properties, expected to provide the basis for a soil fertility index for the study area, are being determined: C org., TN, pH, $CaCO_3$, Exch. Ca, Mg and K and soil particle size.

2.5 Building a Soil Spectral Library

The procedure is based on the method developed by Shepherd and Walsh [4] and includes the following steps: (i) sampling of soil variability within the target area, (ii) measuring of soil spectral reflectance, (iii) sampling of the spectral data space, (iv) acquiring soil attribute data on selected soil samples and (v) calibrating of soil property data to spectra applying multivariate calibration methods.

Air-dried and to 2 mm grinded soil samples were filled into Duran glass Petri-dishes and reflectance spectral readings were measured with a FieldSpec PRO FR spectroradiometer at wavelengths from 350 to 2500 nm at an interval of 1 nm. For standardization the samples were illuminated with an artificial light source from the bottom with a High Intensity Mug Light. In order to reduce differences in light scattering, samples was measured twice at angles differing by 90°.

2.6 Statistical methods

Pre-processing of the raw spectral data prior to statistical analysis was also carried out as described by Shepherd and Walsh [4]. The following steps were carried out: (i) resampling of relative reflectance spectra by selecting every tenth nanometre in order to reduce the volume of data and to match it more closely to the spectral resolution of the instrument, (ii) transformation of the data by first derivative processing using a Savitzky–Golay filter for minimization of variation among samples caused by variation in grinding and (iii) omitting of spectral bands 350 to 380 nm, 970 to 1010 nm and 2460 to 2500 nm, since these bands have low signal to noise ratio or display noise due to splicing between the individual spectrometers [13]. Alternatively to first derivative processing Multiple Scatter Correction (MSC) was tested, but did not improve calibrations.

Selection of samples for chemical analysis was conducted based on principal component analysis on the first derivatives of soil spectral reflectance data using Unscrambler version 7.5 [14]. For each sampling cluster 6-7 samples were selected with regard to an even distribution of the samples over the range of first and second principle components.

Decision tree modelling was carried out using the software CART (Classification And Regression Trees) [15], based on binary recursive partitioning [16]. Decision tree models reveal the hierarchical structures of variables and indicator sets and determine thresholds for classification of data. Decision tree modelling was conducted (i) to build regression models for calibration of TN contents determined by chemical analysis against the 198-reflectance wavebands and (ii) to build classification models for mapping of hotspots and brightspots using information derived from available raster datasets for the plots where groundtruth had been collected. Rules and thresholds determined this way, serve for mapping with knowledge based imagery classification systems.

3 FIRST RESULTS AND DISCUSSION

3.1 Field observations

In a first attempt to gain an overview on the state of the land resources in the study area the data recorded in the field has been assessed focusing on land cover/land use, land use changes as well as water erosion processes. Since a random sampling scheme was applied, it is assumed that the dataset presented here is representative for the situation prevailing in the agricultural land of the test areas. Only the datasets from the test areas Faizabad and Yavan including two thirds of the data collected are being used here.

Land cultivation history has been classified as follows: (i) cultivation for more than 15 years / area being cultivated also during soviet times (ii) cultivation only started after 1992, (iii) cultivation started after 1992 but has been abandoned again and (iv) never cultivated. As can be seen from Table 1, since 1992 considerable land use changes took place: Of the recorded plots only 9 % have been cultivated already during soviet times. Today all in all 27 % of the area is managed, arable area. Additional 12 % of the land has been under cultivation since 1992, but cultivation has now been abandoned again. Therefore it can be assumed that on around 30 % of the total area, formerly used as grazing land (semi-natural and natural areas) ploughing has taken place since 1992.

Table 1. Distribution of the average slope steepness and the share of plots showing visible signs of water erosion with regard to major land cover class and land cultivation history type. The total number of observations is 316.

Major land cover class according to FAO-LCCS	Land Cultivation History Type	Distribution of Land Cultivation History Type recorded [%]	Average slope steepness, standard deviation in brackets [Slope %]	Percentage of plots within the specific LU class showing water erosion [%]
Managed (arable) areas	Cultivated > 15 years	9	14 (8.8)	10
	Cultivated since 1992	18	19 (10.1)	61
	Total	**27**	**18 (10.3)**	**43**
Semi-natural and natural areas	Abandoned	12	22 (11.3)	59
	Never cultivated	61	31 (15.5)	48
	Total	**73**	**29 (15.1)**	**50**

Further more land turned into managed land after 1992 is situated on steep slopes. The figures in Table 1 show that areas still cultivated today are situated on slopes with an average of 19 % steepness, the areas now abandoned on slopes with an average of 22 %. Considering that only in 17 % of the managed land SWC measures such as contour ploughing were observed, it is not surprising that 61 % of the plots cultivated since 1992 and 59 % of the plots now abandoned show visible signs of soil erosion. However also in 48 % of the semi-natural/natural areas that have never been under cultivation visible signs of erosion were recorded, indicating that pressure on the grazing land is also high.

3.2 Applying the soil sensing approach to new conditions

In order to explore relations between the variation in the soil spectra and in soil properties determined in the field or in the laboratory, first principal component (PC1) was plotted against the second (PC2), while the increase in the value of the soil property of interest is displayed by colours (light yellow to black). The principal component analysis was conducted on first derivatives of the soil spectral reflectance data. PC1 explains 52 % and PC2 24 % of the total variance in the spectral data. A thorough analysis for outlier detection should result in even better expressiveness of the principal components.

The left scatter plot in Figure 2 shows relations between principal components and spectral reflection at wavelength 2330 nm. This waveband has been identified as a strong $CaCO_3$ absorption peak [17]. High reflectance in the 2330 nm waveband indicating low $CaCO_3$ contents coincides also with soil samples of a hue value 5 determined by Munsel Colour Code. These soils are of dark red colour and originate from locations where plutonic mother rock is not covered by quaternary loess deposits. The plot shows the presence of systematic variation in the soil spectra that relates to the $CaCO_3$ absorption peak. The strong relationship between the soil spectra and the hue of the soil colour as well as the relationship to $CaCO_3$ content which is also effecting the soil colour is not unexpected, since the wavebands in the visible range are strongly influencing the principle components (apparent by the high loadings of these variables). Further more the 2330 nm waveband is not independent of the principle components, since it is also included in the principle component analysis.

The right scatter plot in Figure 2 shows increasing contents of TN measured in the lab in relation to the first two principle components. The relation for the data plotted here is weak, but thorough outlier detection and data transformations should improve the relationship.

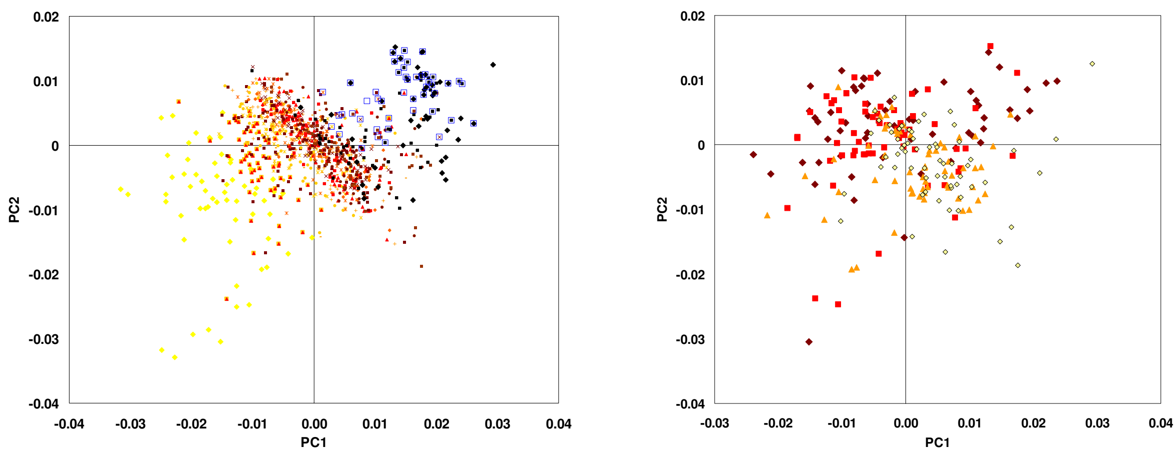

Figure 2. Scatter plots showing relations between variation in soil spectra and soil properties: variations in wavelength 2330nm, being a high $CaCO_3$ absorption peak, and hue of Munsel Colour Code (spots marked with a blue square) for 1050 samples on the left and measured contents of TN [%] for 250 soil samples on the right. Light marks indicating high and dark marks low contents of $CaCO_3$ and TN respectively.

First calibration models for prediction of soil properties to soil spectra have been elaborated for TN using the binary recursive partitioning method implemented in the software CART [15]. The scatter plot in Figure 3 shows actual and predicted values for TN calibration and validation datasets. The respective coefficients of determination are r^2=0.92 and r^2=0.71. 186 samples served for calibration, and 80 randomly selected holdout samples for validation.

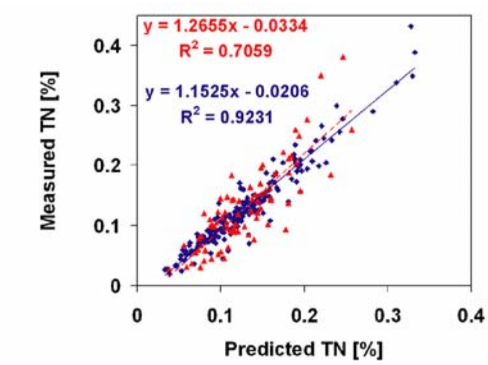

Figure 3. Calibration of soil properties to first derivative reflectance spectra using binary recursive partitioning modelling: Predicted values for TN plotted against measured TN contents: calibration (blue) and validation data (red).

As Shepherd and Walsh [4] explain, such level of prediction accuracy is sufficiently high when spatial and temporal variability of the soil property is large relative to the accuracy of its measurement. Moreover with regard to the inaccuracy of analytical results from Tajik laboratories today, it can be expected that the soil library approach will provide a reliable opportunity for accurate determination of soil properties over the long run.

3.3 Decision trees revealing rules for water erosion mapping

Characteristics of water erosion processes differ distinctively between semi-natural/natural and managed land cover classes. Therefore models for predicting areas with water erosion and without erosion have been run separately for the two types of land cover/land use classes. Several sets of splitting rules are available with CART. The Gini method typically produces best results [16], which was also true for the modelling conducted in this study. Prediction success was determined by 10 fold cross-validation. For the test data of the model presented in Figure 4 it is 61 % and 72 % for the cases no erosion and erosion respectively. Model input data consisted of (i) topographic information extracted from the DEM (slope, aspect, and curvature) (ii) Landsat 7 spectral information (band 1-5) of scenes dating from May 2002 and August 2000 and (iii) the vegetation index OSAVI calculated from these images. In addition to the raster information, the first principle components (PC1+2) calculated for the soil spectral reflectance data of topsoil samples, so far only available as point data, also fed into this model. From the 17 variables available, five layers were found most effective in splitting the data into nodes where cases of “erosion” or “no erosion” prevailed. Interesting that four of those layers contain mainly information on the presence and state of vegetation. The absence of topographic information supports the assumption that it is the land cover/land use characteristics that are decisive for the occurrence of soil erosion processes. The presence of the variable PC1+2 in this model suggests that soil spectral information is relating to the presence of water erosion in a specific area.

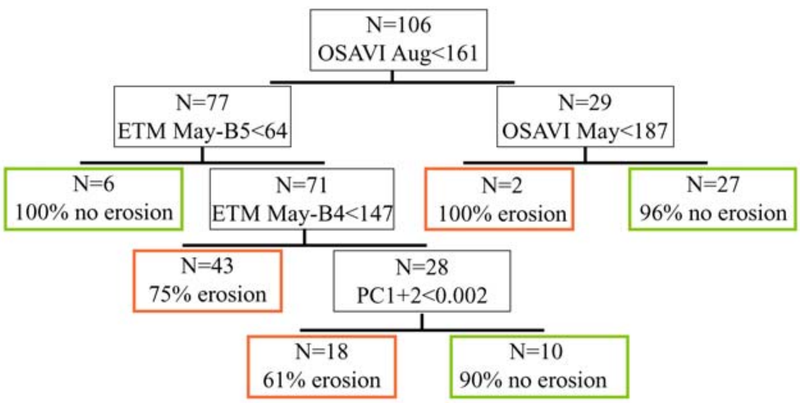

Figure 4. Tree model for mapping water erosion on cultivated areas (erosion or no erosion visible). Groundtruth was collected on 106 plots (N=number of plots). The abbreviations for input data used are as follows. Landsat 7 scene from May 2002 (ETM May), from August 2000 (ETM Aug), bands 4,5 (B4, B5), OSAVI (OSAVI), first and second principal components of soil spectral data (PC1+2). Percentages given in the end nodes represent the number of cases observed with regard to the total number of cases in the respective end node.

4 CONCLUSIONS

The groundtruth data shows that visible signs of water erosion can be observed in 50 % of the plots recorded. It further indicates that the risk for water erosion processes going on is high in areas that only after 1992 have been turned from grazing land into cultivated areas and are situated on rather steep slopes. It has been shown that variety in soil spectral reflectance data is relating to soil properties important for determination of soil fertility in the region, such as $CaCO_3$. Further a first model allowing to predict TN from soil spectral data shows a coefficient of determination of r^2=0.71, which can be considered a sufficient level of accuracy. These results are encouraging for providing efficient, stable and accurate methods for soil property determination in Tajikistan. First results from decision tree modelling for areas with visible signs of water erosion show that an index representing soil properties from spectra will be very helpful in mapping land degradation, especially so if such an index can be calibrated to Landsat 7 satellite imagery data as successfully done by Vagen et.al. [5].

ACKNOWLEDGMENTS

The Soil Science Research Institute of the Tajik Academy of Agricultural Science has and is contributing to the study in scientific, administrative and organizational questions. The Remote Sensing Laboratory (RSL), Institute of Geography of the University of Zurich, provided the spectroradiometer. The investigations were supported by the National Centre for Competence in Research North-South (NCCR) / Individual Project 'Natural Resources and Ecology' (IP2), financed by the Swiss National Science Foundation (SNSF) and the Swiss Agency for Development and Cooperation (SDC).

REFERENCES

[1] ADB, 2001: Agricultural Assessment Project (ASAP) Tajikistan. TA 3295 – TAJ. Revised Final Report, pp.48

[2] SADIKOV, KH.R. (ed.), 1999: Abstracts of the International Conference on Mountain regions of Central Asia. Sustainable Development Issues. Dushanbe, Tajikistan.

[3] ERGASHEV, M., WOLFGRAMM, B. AND NEKUSHOEVA, G., 2005: Individual Approach for Improving the Productivity of Grazing Land. In: Critchley W., Liniger H.P. (eds.): Local responses to global land degradation. (Forthcoming).

[4] SHEPHERD, K. AND WALSH, M., 2002: Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Sci. Soc. Am. J. 66*, pp. 988–998.

[5] VÅGEN, T.-G., SHEPHERD, K. AND WALSH, M., 2004: Vis-Nir spectroscopy for characterisation of landscape level change in soil quality following deforestation and conversion in the highlands of Madagascar. *Geoderma* (submitted).

[6] KUTEMINSKIJ, V.J., LEONTEVA, R.S., 1966: Soils of Tajikistan [In Russian]. Dushanbe: Tajik Soil Science Institute, Ministry of Agriculture of Tajikistan.

[7] RICHTER, R., 2005: Atmospheric / topographic correction for satellite imagery", DLR IB 565-01/05, Wessling, Germany.

[8] RONDEAUX, G., STEVEN, M. AND BARET, F., 1996: Optimisation of soil-adjusted vegetation indices. *Remote Sensing of Environment, 55*, pp. 95–107.

[9] WALSH, M., 2004: Baseline and Project Monitoring Plan for the Western Kenya Integrated Ecosystem Management. Paper in progress, World Agroforestry Centre (ICRAF), Nairobi, Kenya.

[10] LINIGER, H.P. AND SCHWILCH, G., 2002: Better decision making based on local knowledge - WOCAT method for stainable soil and water management. *Mountain Research and Development Journal, 22 Vol. 1*

[11] LYNDEN VAN, G.W.J., LINIGER, H.P. AND SCHWILCH, G., 2002: The WOCAT mapping methodology, a standardized tool for mapping degradation and conservation. *Proceedings of ISCO Conference 2002*, Beijing.

[12] DI GREGORIO, A., AND JANSEN, L.J.M., 1998: Land Cover Classification System (LCCS): Classification Concepts and User Manual. FAO, Rome.

[13] ANALYTICAL SPECTRAL DEVICES INC., 1997: FieldSpecTM User's guide. Analytical Spectral Devices Inc., Boulder, CO.

[14] CAMO Inc., 1998: The Unscrambler user manual. CAMO Inc., Corvallis, OR.

[15] Breiman, L., Friedman, J., Olshen, R. and Stone, Ch., 1984: Classification and Regression Trees. Wadsworth, Pacific Grove.

[16] STEINBERG, D. AND COLLA, P., 1995: CART: Tree-Structured Non-Parametric Data Analysis. Salford Systems, San Diego, CA.

[17] GAFFEY, F.Y., 1986: Spectral reflectance of carbonate minerals in the visible and in near infrared (0.35-2.55um): calcite, aragonite and dolomite. *American Minerals 71*, pp. 151-162.